

## **GPLSI Sentiment Analysis v1.0: Análisis de sentimientos en textos**

### **GPLSI Sentiment Analysis v1.0: Sentiment analysis on texts**

**Resumen de la aplicación:** GPLSI Sentiment Analysis v1.0 constituye una interfaz de programación de aplicaciones (API) que incluye librerías de programación útiles para sistemas de terceros. El objetivo de esta tecnología es analizar textos y detectar la polaridad de los sentimientos (estados de opinión) expresados en ellos. La funcionalidad para detectar la polaridad de los sentimientos implica dos niveles de granularidad: polaridad a nivel global (PNG) y polaridad a nivel de aspectos (PNA). La PNG del texto se define como el análisis de todo el texto para determinar el estado de opinión (positivo, negativo o neutral). La PNA determina el estado de opinión para un aspecto en particular de una entidad dentro un contexto determinado. Por ejemplo, en un mismo texto (usualmente comentarios de usuarios) podemos encontrar opiniones diferentes y opuestas sobre los actores de una película y su banda sonora. Ambas funcionalidades, PNG y PNA, se basan en clasificadores SVM (Support Vector Machines) que son alimentados con patrones léxicos de skip-grams obtenidos de textos de entrenamiento. En el caso de PNG, estos clasificadores han sido entrenados sobre los conjuntos de datos de la competición Semeval task 2 (10.709 comentarios de usuarios) y Metacritic (5.888 críticas de usuarios sobre películas). Con respecto a PNA, se ha utilizado como conjunto de entrenamiento el corpus de la competición Semeval 2014 - Task 4 - Subtask - 1 - Restaurants. Los resultados para PNG oscilan entre el 62% y el 88% de precisión, dependiendo de la calidad de redacción del texto analizado. Con respecto a PNA, los resultados se encuentran alrededor del 61% de precisión. Una explicación detallada de ambos algoritmos se puede ver en [GTF15].

**Summary:** Summary: GPLSI Sentiment Analysis v1.0 is an application programming interface (API) which involves programming libraries for third-parties. The aim of this technology is to analyse texts to detect sentiment polarities. Two different levels of analysis are defined in this tool: global polarity (GP) and aspect-based polarity (ABP). GP allows inferring the sentiment polarity expressed in the whole text, whereas ABP determines the polarity regarding particular aspect mentioned in it. Both functionalities are based on Support Vector Machines. The GP algorithm have been trained by using the following datasets: Semeval task 2 datasets (10,709 user comments) and Metacritic (5,888 user reviews about movies); regarding ABP, the training dataset has been Semeval 2014 - Task 4 - Subtask 1 – Restaurants. The performance achieved by GP ranges between 62% to 88% precision. The changes in performance depends on the quality of the style of the text. A detailed explanation about the algorithms, evaluation, and analysis can be found in [GTF15].

# Especificaciones técnicas/ Technical Specifications

Lenguaje de programación/ Development Language: Java

Entorno Operativo/ Operating Environment: Linux / Windows

Versión/Version: 1.0

## Estructura de ficheros/Folder structure:

### Directorio/Folder ->

#### sentiment-analysis-core:

- nbactions-giventext.xml
- nbactions.xml
- pom.xml
- nbactions-multitextscommand.xml
- nb-configuration.xml

#### Directorio/ Folder ->

##### sentiment-analysis-core/src/main/resources/ models:

- en-chunker.bin
- en-ner-date.bin
- en-ner-money.bin
- en-ner-percentage.bin
- en-ner-time.bin
- en-pos-maxent.bin
- en-sent.bin
- testmallet.txt
- en-ner-asset.bin
- en-ner-location.bin
- en-ner-organization.bin
- en-ner-person.bin
- en-parser-chunking.bin
- en-pos-perceptron.bin
- en-token.bin

#### Directorio/ Folder ->

##### sentiment-analysis-core/src/main/resources/t raining:

- polarity.data

#### Directorio / Folder ->

##### sentiment-analysis-core/src/main/java/es/ua/ sentimentanalysis:

- Main.java

### Directorio / Folder ->

#### sentiment-analysis-web:

- nb-configuration.xml
- pom.xml

#### Directorio / Folder ->

##### sentiment-analysis-web/src/main/java/es/ua/

- Subdirectorio / Subfolder -> analyzer

- Definition.java
- MorphoSyntacticAnalysis.java
- POSTagger.java
- TextEntity.java
- SentimentAnalyzer.java

- Subdirectorio / Subfolder ->

##### contextsimilarity

- ContextSimilarityFactory.java
- ContextSimilarityStrategy.java

- Subdirectorio / Subfolder ->

##### sentiment

- EmotionAttributes.java
- EmotionDetector.java
- EmotionReader.java
- SentimentAnalysis.java
- SentimentAttributesEmotionS  
cores.java
- SentimentAttributes.java
- SentimentAttributesSocial.jav  
a
- SentimentAttributesSocialSco  
re.java
- SentimentSocialAttributesEm  
otionScoresResponse.java
- SentimentSocialAttributesRes  
ponse.java

- Subdirectorio / Subfolder -> util:

- GsonInstance.java
- LookUpEntity.java
- PorterStemmer.java
- LevenshteinCalculator.java
- Pair.java

#### sentimentanalysis/application:

- ApplicationConfig.java

#### Directorio / Folder ->

##### sentiment-analysis-web/src/main/java/es/ua/

##### sentimentanalysis/services:

- SentimentAnalysis.java
- Subdirectorio / Subfolder ->

#### exception

- InvalidParametersException.java
- **Subdirectorio / Subfolder -> filter**
  - EntityFinderFilter.java
- **Subdirectorio / Subfolder -> internal/annotation**
  - ApplicationInfo.java
- **Subdirectorio / Subfolder -> internal/utills**
  - AnnotatedTextProcessor.java
  - ReflectionUtils.java
- **Subdirectorio / Subfolder -> listener**
  - ContextListener.java
- **Subdirectorio / Subfolder -> parameters**
  - IParameters.java
  - AnaysisSentimentParameters.java
- **Subdirectorio / Subfolder -> security**
  - SecurityUtils.java
  - UntrustedAccessExceptionMapper.java
  - UntrustedAccessException.java

#### Directorio / Folder ->

##### sentiment-analysis-web/src/main/resources/models:

- en-chunker.bin
- en-ner-date.bin
- en-ner-money.bin
- en-ner-percentage.bin
- en-ner-time.bin
- en-pos-maxent.bin
- en-sent.bin
- testmallet.txt
- en-ner-asset.bin
- en-ner-location.bin
- en-ner-organization.bin
- en-ner-person.bin
- en-parser-chunking.bin
- en-pos-perceptron.bin
- en-token.bin

#### Directorio / Folder ->

##### sentiment-analysis-web/src/main/resources/training:

- polarity.data

#### Directorio / Folder ->

##### sentiment-analysis-web/src/main/webapp:

- entityfinder.html
- index.html
- logs.html
- **Subdirectorio / Subfolder -> css**
  - bootstrap.min.css
  - bootstrap-theme.min.css
  - log-style.css
  - main.css
  - style.css

#### ● Subdirectorio / Subfolder -> doc

- **Subdirectorio css**
  - reset.css
  - screen.css
- **Subdirectorio / Subfolder -> img**
  - bg.png
  - logo.png
  - pet\_store\_api.png
  - wordnik\_api.png
  - explorer\_icons.png
  - logo\_small.png
  - throbber.gif
- index.html.bak
- o2c.html
- swagger-ui.min.js
- favicon.ico
- index.html
- **Subdirectorio / Subfolder -> lib**
  - backbone-min.js
  - jquery.slideto.min.js
  - swagger.js
  - handlebars-1.0.0.js
  - jquery.wiggle.min.js
  - swagger-oauth.js
  - handlebars-2.0.0.js
  - marked.js
  - underscore-min.js
  - highlight.7.3.pack.js
  - **Subdirectorio shred**
    - content.js
  - underscore-min.map
  - jquery-1.8.0.min.js
  - shred.bundle.js
  - jquery.ba-bbq.min.js
  - swagger-client.js
  - swagger-ui.js
- **Subdirectorio / Subfolder -> fonts**
  - glyphicons-halflings-regular.eot
  - glyphicons-halflings-regular.ttf
  - glyphicons-halflings-regular.svg
  - glyphicons-halflings-regular.woff
- **Subdirectorio / Subfolder -> img**
  - bg-header.png
  - bg.png
  - gplsi.png
- **Subdirectorio / Subfolder -> js**
  - ajax.js
  - bootstrap.min.js
  - jquery-2.1.1.min.js
  - logs.js
  - app.js
  - **Subdirectorio controllers**

- main.js
  - jquery-ui.js
- **Subdirectorio / Subfolder -> META-INF**
  - context.xml
- **Subdirectorio / Subfolder -> views**
  - main.html
  - **Subdirectorio / Subfolder -> fragments**
- footer.html
  - header.html
- **Subdirectorio / Subfolder -> WEB-INF**
  - host.properties
  - log4j.properties
  - security.properties
  - test.txt
  - web.xml

**Requerimientos:** Java 1.7 o superior instalado, 3 GB de RAM o superior, 1 GB de disco duro libre en sistema, Apache Tomcat 8.0 o superior y Maven. Además, son necesarios los modelos para realizar el análisis de sentimientos (i.e. “**polarity.data**”), así como los modelos de Apache OpenNLP<sup>1</sup>. Los modelos OpenNLP ya se incluyen en la distribución de la aplicación, por comodidad del usuario, pero no necesariamente se tiene que distribuir dentro de ésta, ya que son recursos externos.

**Instalación:** Descargar el código fuente del repositorio de control de versiones. Se deben generar los binarios tanto de la aplicación de consola como web utilizando Maven. Antes de generar los binarios es importante asegurarse de que los modelos de análisis de sentimiento así como los de Apache OpenNLP se encuentran en los directorios [core|web]/src/main/resources/training y [core|web]/src/main/resources/models, respectivamente.

En ese momento la aplicación de consola está lista para funcionar.

La instalación de la aplicación web es opcional y, además de lo anterior, es necesario desplegar el war generado en el servidor de aplicaciones Tomcat.

En principio se puede usar cualquier sistema operativo, pero sólo ha sido probado en Ubuntu<sup>2</sup>.

**Ejecución:** La aplicación web dispone de una página de documentación del servicio web REST. El servicio se puede ejecutar directamente desde esta documentación en un navegador o mediante el uso clientes para el servicio web REST.

La aplicación de consola dispone de un comando para detectar la polaridad de los sentimientos expresados en textos en idioma inglés (a nivel global y de aspectos), así como un comando de ayuda. A continuación mostramos algunos ejemplos:

- **[\$[nombre del jar] help analyze**
  - Muestra la ayuda del comando, con los parámetros y su descripción.
- **[\$[nombre del jar] analyze -f/--file [nombre del archivo txt]**
  - Permite proporcionar un archivo con el texto que se desea analizar. Por defecto, el análisis de polaridad se realiza solo a nivel global. Cada texto debe

<sup>1</sup> Disponibles en <http://opennlp.sourceforge.net/models-1.5/>

<sup>2</sup> <https://www.ubuntu.com/>

ser un párrafo dentro del archivo, separado por un salto de línea. Para más información consultar la ayuda, ver comando anterior.

### Dependencias de programas de terceros:

El listado completo de dependencias, tanto para la aplicación de consola como para el servicio web, se encuentra en los archivos pom.xml. Aquí se incluye un listado de las dependencias que se descargarán de varios repositorios maven, en el cual las librerías propias del GPLSI se marcan con con un \* y en negrita:

#### sentiment-analysis-core 1.0.0

- airline 0.7
- commons-lang 3-3.1
- gson 2.3
- guava 19.0
- jettison 1.3.3
- junit 4.12
- lucene-analyzers-common 4.9.0
- lucene-queryparser 4.9.0
- opennlp tools 1.5.3
- **sam-sentiment-core 1.1.0\***
- annotations 2.0.3
- **classifier-base 1.0.0\***
- commons-io 2.4
- hamcrest-core 1.3
- jackson-annotations 2.5.0
- jackson-core 2.5.0
- jackson-databind 2.5.0
- java-cup 0.11a
- javax.inject 1
- jwnl 1.3.3
- lucene-core 4.9.0
- lucene-queries 4.9.0
- lucene-sandbox 4.9.0
- opennlp maxent 3.0.3
- slf4j-api 1.7.21
- slf4j-api 1.0.1
- stax-api 1.0.1
- weka-stable 3.6.14

#### Web

- **sentiment-analysis-core 1.0.0\***
- gson 2.3
- guava 15.0
- jackson-annotations 2.5.0
- jackson-core 2.5.0
- jackson-databind 2.5.0
- jackson-jaxrs-base 2.5.0
- jackson-jaxrs-json-provider 2.5.0
- jackson-jaxrs-xml-provider 2.5.0
- jackson-module-jaxb-annotations 2.5.0
- javaxee-web-api 7.0
- javassist 3.18.1-GA
- javax.annotation-api 1.2
- javax.servlet-api 3.0.1
- javax.ws.rs-api 2.0.1
- jaxb-api 2.2.7
- jersey-client 2.15
- jersey-common 2.15
- jersey-container-servlet 2.15
- jersey-container-servlet-core 2.15
- jersey-entity-filtering 2.15
- jersey-guava 2.15
- jersey-media-json-jackson 2.15
- jersey-media-multipart 2.15
- jersey-server 2.15
- log4j 1.2.17
- persistence-api 1.0
- swagger-annotations 1.3.11
- swagger-core\_2.10-1.3.11
- swagger-core 1.5.3
- swagger-jaxrs\_2.10-1.3.11
- swagger-jersey2-jaxrs\_2.10-1.3.11
- swagger-utils\_2.10-1.3.11
- swagger-models 1.5.3
- validation-api 1.1.0
- xom 1.2.10
- airline 0.7
- annotations 2.0.3
- aopalliance-repackaged 2.4.0-b06
- **classifier-base 1.0.0\***
- commons-io 2.5
- commons-lang3-3.1
- commons-lang 2.4
- dom4j 1.6.1
- hamcrest-core 1.3

- hk2-api 2.4.0-b06
- hk2-locator 2.4.0-b06
- hk2.hk2-utils 2.4.0-b06
- jackson-dataformat-xml 2.5.0
- jackson-module-jsonSchema 2.1.0
- jackson-module-scala 2.10-2.41
- java-cup 0.11a
- javax.inject 2.4.0-b06
- javax.inject 1
- joda-convert 1.2
- joda-time 2.2
- json4s-ast\_2.10 3.2.9
- json4s.json4s-core\_2.10 3.2.9
- json4s.json4s-ext\_2.10 3.2.9
- json4s.json4s-jackson\_2.10 3.2.9
- json4s.json4s-native\_2.10 3.2.9
- jsr305 2.0.1
- jsr311-api 1.1.1
- junit 4.12
- jwnl 1.3.3
- lucene-core 4.9.0
- lucene-queries 4.9.0
- lucene-queryparser 4.9.0
- lucene-analyzers-common 4.9.0
- lucene-sandbox 4.9.0
- mimepull 1.9.3
- opennlp-maxent 3.0.3
- opennlp-tools 1.5.3
- osgi-resource-locator 1.0.1
- paranamer 2.6
- reflections 0.9.9-RC1
- **sam-sentiment-core 1.1.0\***
- scala-compiler 2.10.0
- scala-library 2.10.0
- scala-reflect 2.10.4
- scalap 2.10.0
- slf4j-api 1.6.3
- stax-api 1.0.2
- stax2-api 3.1.1
- swagger-annotations 1.5.3
- woodstox-core-asl-4.4.0
- xalan 2.7.0
- xercesImpl 2.8.0
- xml-apis 1.3.03
- weka-stable-3.6.14

## Referencias/References

[GTF15] Gutiérrez, Y.; Tomás, D.; Fernández, J. Benefits of Using Ranking Skip-Gram Techniques for Opinion Mining Approaches. Proceedings of eChallenges 2015 e-2015. pp 1-10. 2015